Open Access in High Energy Physics: Overview and Future Challenges

Martin Köhler

DESY, Notketraße 85, 22607 Hamburg, Germany

DOI: http://dx.doi.org/10.3204/DESY-PROC-2008-02/koehler_martin

This paper describes a business model for the 'golden' road to Open Access which is specifically tailored for High Energy Physics (HEP), but can possibly be extended to other disciplines: SCOAP^3 - the **S**ponsoring **C**onsortium for **O**pen **A**ccess **P**ublishing in **P**article **P**hysics. We also report on the status of INSPIRE, the designated successor of the well-established SPIRES database, and a complete HEP e-Infrastructure. Finally, aspects of the challenges involved in HEP data preservation are discussed.

1 Introduction to Open Access

According to Wikipedia "*Open Access* (OA) is free, immediate, permanent, full-text, online access, for any user, web-wide, to digital scientific and scholarly material." The immediate and permanent access to the full-text of (research) articles for anyone webwide is made possible by consent of the author and/or the copyright holder. One should note that the concept of OA is compatible with peer review and publishing in non OA-journals. Furthermore, OA is not free to produce and there is much debate about the economics of its funding. In general, specific business models depend on the way OA is delivered (see below). The most prominent roads to OA are *Open Access publishing* - the 'golden' road - and *Open Access self-archiving* - the 'green' road.

2 Open Access and the role of journals in the High Energy Physics community

The idea of Open Access is well known and long established in the HEP community. For over forty years, HEP institutes have shipped preprints of their authors worldwide at their own expenses. In 1991, the HEP community saw the launch of arXiv –http://xxx.lanl.gov at that time– the archetype of Open Archives. With the establishment of arXiv, OA became second nature to HEP scientists and nowadays posting on arXiv before even submitting to a journal is quite a common practice. It is worth mentioning that this process has always been purely author driven, without mandate, even without debate.

As an 'all-arXiv discipline' and due to spiralling subscription costs HEP is at a risk to see numerous journal subscriptions cancelled by large multidisciplinary university libraries (if that is not already the case). On the other hand, high quality peer-reviewed journals are the basis

for academic evaluation of institutes and (young) researchers. The main role of journals is to assure high-quality peer-review and act as keepers-of-the-records.

Since OA obviously is incompatible with the traditional subscription model, publishers and customers have experimented with new OA compatible business models. In the *hybrid model*, for an additional fee in the range from $\approx$ 900 – 3.000 US$, it is possible to make individual articles openly accessible. The subscription costs might get reduced, but for institutions and libraries, the hybrid model is generally more expensive than the traditional subscription model and hence has had very little success up to now. In the *sponsoring model* institutions pay for the journals through annual sponsoring fees; there are no author charges and the whole journal content is open access. Another prominent model, which is very successful in life science is the *author charge* model. Here a publishing fee is due for all accepted articles and again the whole journal is then open access.

3 The SCOAP^3 consortium

In 2008 the Large Hadron Collider (LHC) at CERN, will begin to operate. It will be the only HEP experiment worldwide of this size and most HEP publications will be related to this machine and the results of the experiments done with it. It is therefore very significant that in 2007 all the LHC collaborations, i.e. more than 5400 scientists, have signed a statement strongly encouraging the usage of electronic publishing methods and supporting the principle of Open Access. To support this goal CERN and other leading HEP laboratories, DESY among them, have started SCOAP^3, - the **S**ponsoring **C**onsortium for **O**pen **A**ccess **P**ublishing in **P**article **P**hysics. SCOAP^3 plans to make the approx. 20,000 articles/year published by the HEP community freely accessible by re-directing subscription money on a budget of approx. 10M€/year. This money will be raised from about 50 funding bodies and roughly 10 publishers will be involved[1]. Note that this scale is much smaller than that of e.g. the currently well established ATLAS HEP experiment, where the budget is approx. 400M€ and more than 1000 contracts have been signed. At the moment about 38% of the needed funds have been pledged and another 15% are coming soon. The substantial goal is to have SCOAP^3 operational for the first LHC articles.

4 HEP databases

A recent poll[2] of the HEP community with more than 2.000 participants, which is roughly 10% of the whole community has demonstrated that more than 90% of the users use community-based services, e.g. subject repositories and laboratory-supported databases. The use of Google and Google scholar correlates with experience. While almost 22% of the younger people with less than two year career experience use these database for information recovery, this drops to 6% for people with more than six years of experience. The use of (other) commercial databases is marginal. With almost 50%, the well-established SPIRES database is by far the largest of the laboratory-supported databases and was one of the first on the 'market'. It has been initially built by SLAC and is today maintained in collaboration by SLAC, DESY and FERMILAB.

While the quality of the data contained in SPIRES is undoubtedly extremely high, the user-interface and the underlying software architecture is due for a major update. Moreover, according to the above-mentioned poll, the scientists obviously do not only need a modern user-interface for SPIRES, but also a tighter integration of the various different HEP information

systems and new features, i.e. they need a new HEP e-Infrastructure, which includes text- and data-mining applications, as well as Web 2.0 technology. To achieve this goal, in Spring 2008, the SPIRES collaboration and CERN have signed an expression of interest to build such a system, called INSPIRE, with CERN's Invenio as the underlying software. The first alpha test of the new system is due in autumn 2008, a tighter collaboration with publishers and other data providers will follow soon, and public beta tests will start in 2009.

5 The next frontier: Preservation of research data

During the last decades, the HEP experiments became more and more complex and more expensive. Given the fact that e.g. the LHC has cost about 6 billion Euros in public money, it is obvious that one can not repeat these experiments later to obtain new or validate the old data, put new theories to test or to combine the data with some future experiments. The only feasible option is storage and re-use of the collected data. In fact, if this data cannot re-used after the experiment has been stopped and disbanded, the investment into the experiment will not have been exploited to its full capacity. Only an additional and relatively small fraction of the funds is needed to preserve a large fraction of knowledge. CERN Director General Elect Rolf-Dieter Heuer suggests that in order to preserve the HEP data the community should follow a *parallel way*: In addition to experiment data models, a parallel format for (re-)usable high-level objects is to be elaborated. This approach has successfully worked in the past to combine data of 'competing' experiments. 'Oral' and 'additional' knowledge is embedded into such a format to make it understandable and hence re-usable by practitioners in other experiments as well as by theorists.

Unfortunately, there are issues with the 'parallel way': First of all even a small fraction of a big number still gives a large number, so substantial funds are needed. Elaboration of a parallel format competes with research time and the thousands of person-years needed for the 'parallel way' need enormous academic incentives and motivation for realization. This can only succeed if the issues of (open) access, credibility, accountability, reproducibility of results and depth of peer reviewing have been addressed.

To summarize: A monolithic way of doing business needs rethinking.

6 Conclusion

After more than forty years of preprints, sixteen years of repositories and the web, SCOAP^3 –a model for Open Access Publishing– is the next logical step in HEP publishing. On the information retrieval side, the time is ripe for INSPIRE, an e-Infrastructure for HEP communication. The next challenge is the preservation of HEP data.

References

[1] S. Mele *et al.*,J. High Energy Phys. **12** S01 (2006).

[2] A. Gentil-Beccot*et al.* arXiv:0804.2701 [cs.DL].